



REDES NEURAIS COM TENSORFLOW

DIEGO RODRIGUES DSC
INFNET

Agenda

Parte 1 : Meta Heurística de Treinamento Robusta

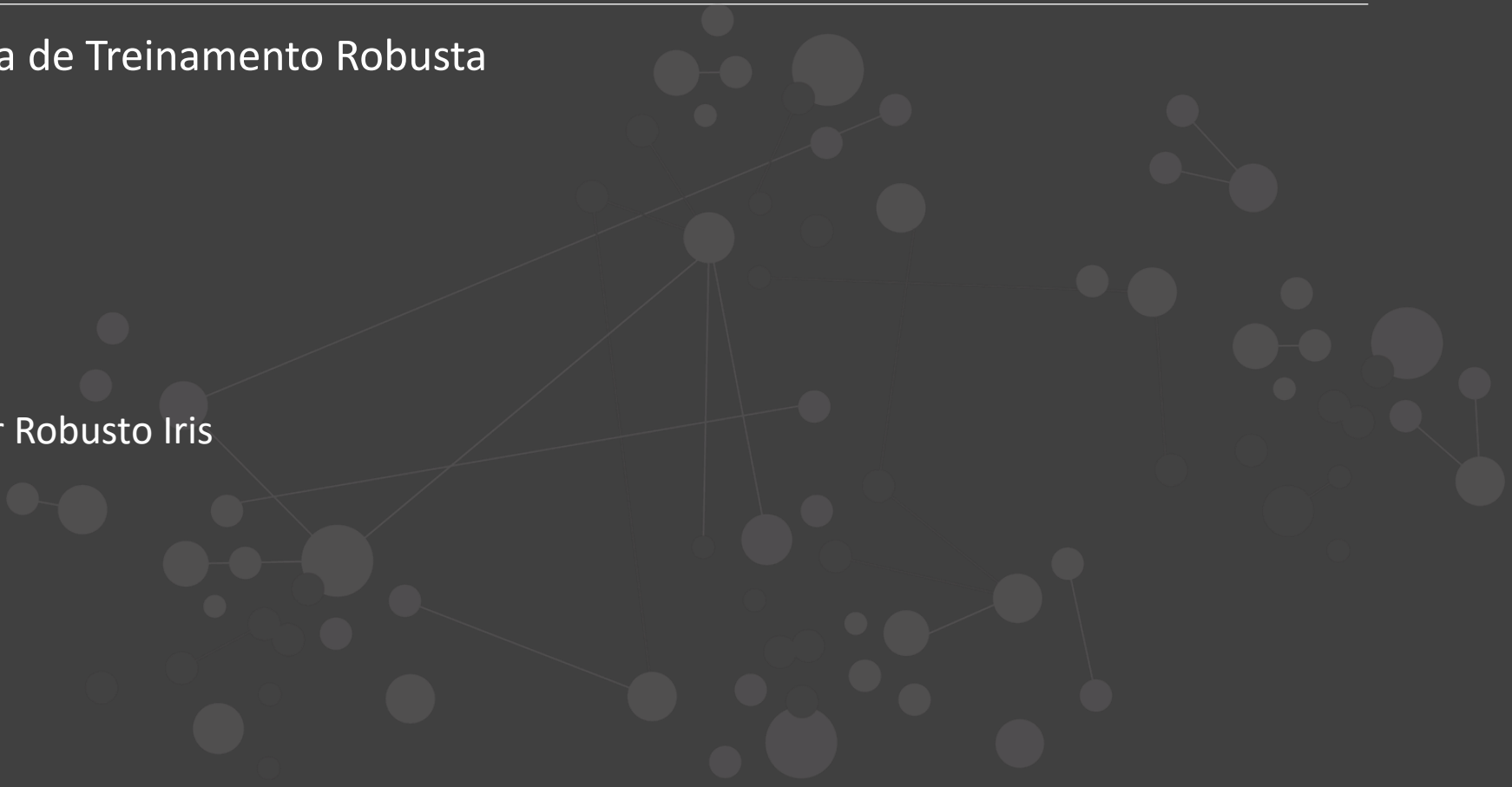
- Novo ciclo do CRISP
- Representação
- Modelagem
- Validação

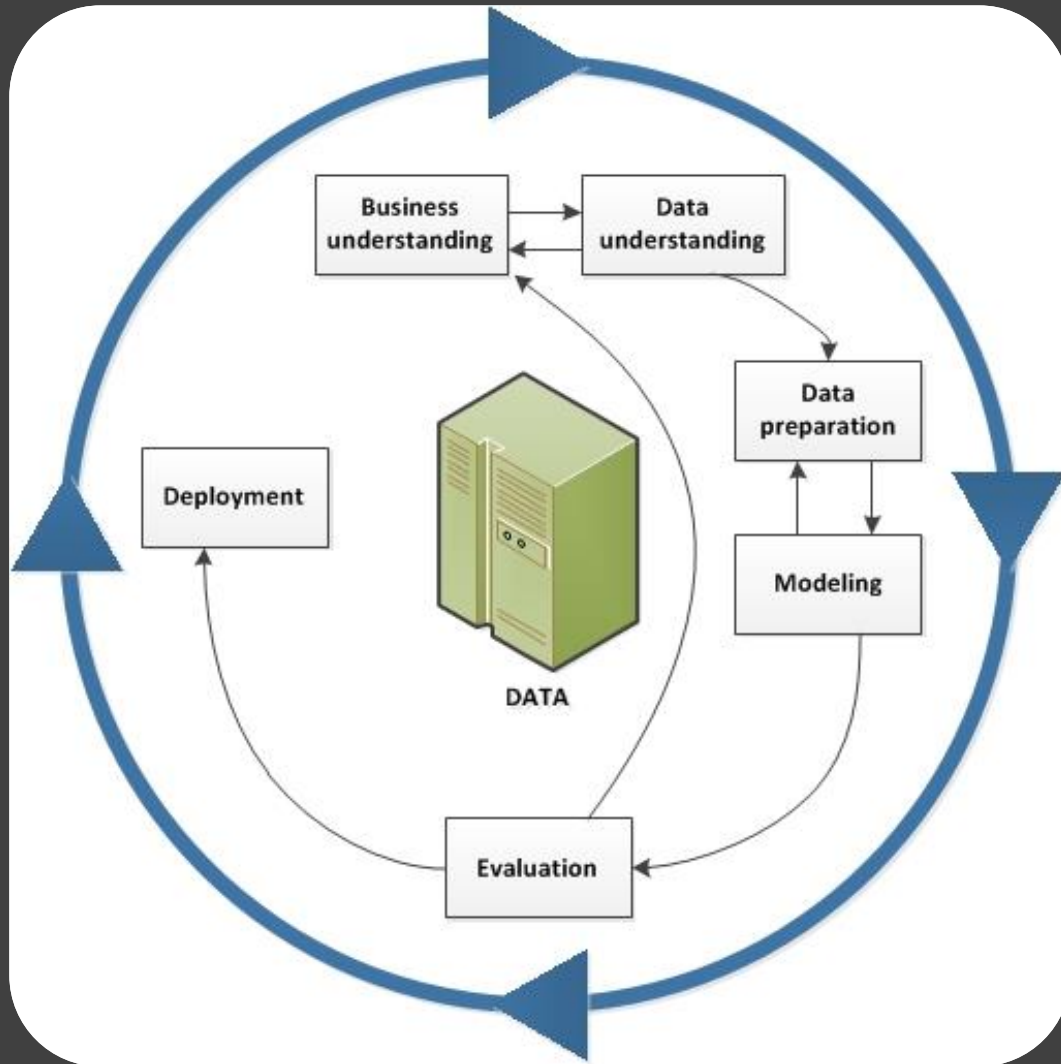
Parte 2 : Prática

- Notebook: Classificador Robusto Iris

Parte 3 : Trabalhos

- Escopo & Evolução





Cross
Industry
Process for
Data Mining
(CRISP-DM)

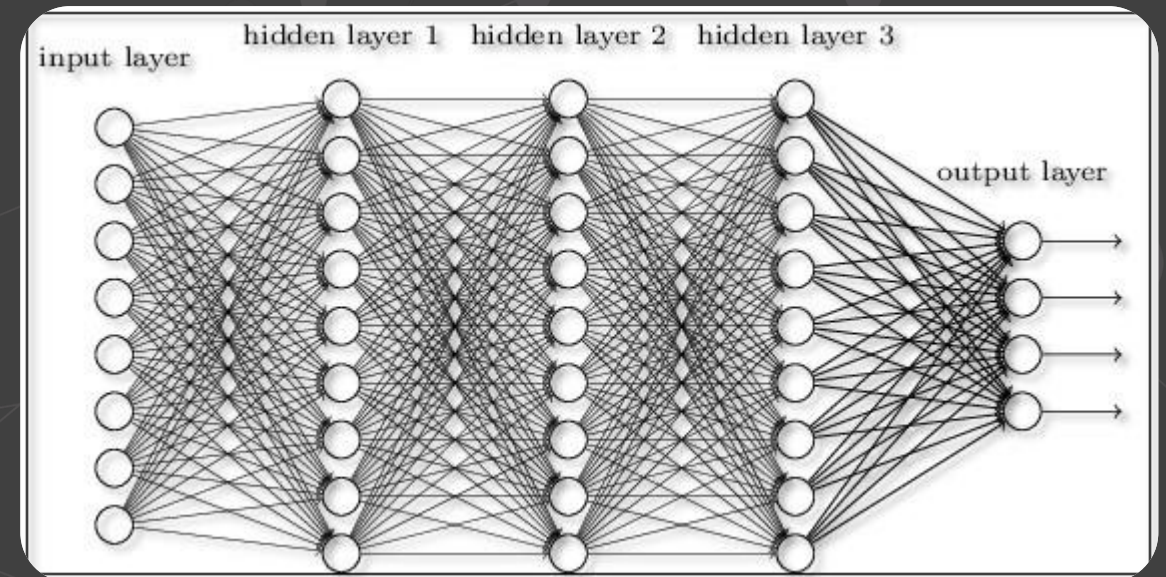
Novo Ciclo CRISP

Algoritmo	Representação	Preparação	Modelagem	Validação
<ul style="list-style-type: none">• Reta 2 Pontos• NN 10% VAL• NN 10 Folds	<ul style="list-style-type: none">• 2D• 2D• 2D	<ul style="list-style-type: none">• Nenhuma• Nenhuma• Scale	<ul style="list-style-type: none">• Reta 2 Pontos• NN Básica• NN Hidden	<ul style="list-style-type: none">• Nenhuma• Precisão/Recall• Precisão/Recall

- Garantir estabilidade no treinamento
- Chegar a mesma solução independente do experimento
- Garantir Generalização

Análise de Negócio

- **Reprodutibilidade do Experimento**
 - **Controlar SEED do Numpy & Keras.**
 - **Mitigar o efeito da inicialização dos Parâmetros**

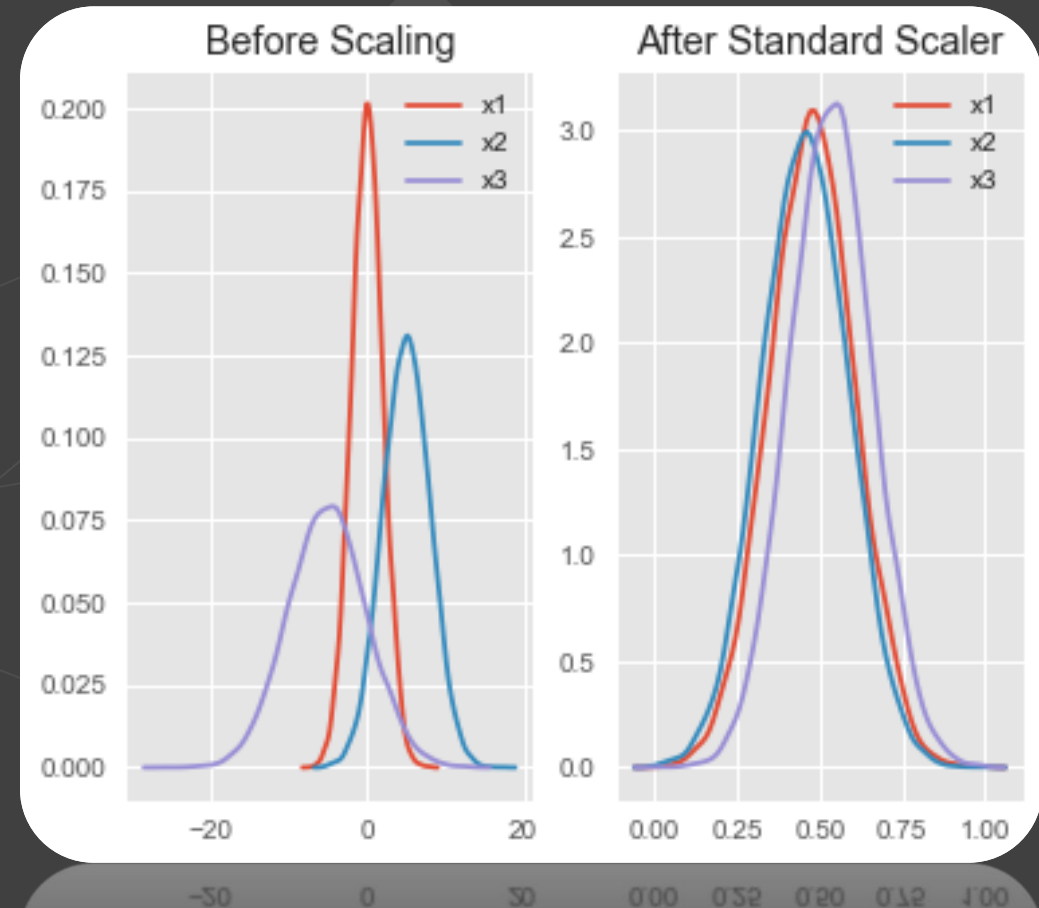


Representação

- **Normalização**
 - Garantir que as variáveis independentes possuem a mesma escala
- **Análise de Componentes Principais**
 - Garantir que as variáveis independentes sejam descorrelacionadas
- **Análise de Componentes Independentes**
 - Garantir que as variáveis independentes sejam independentes

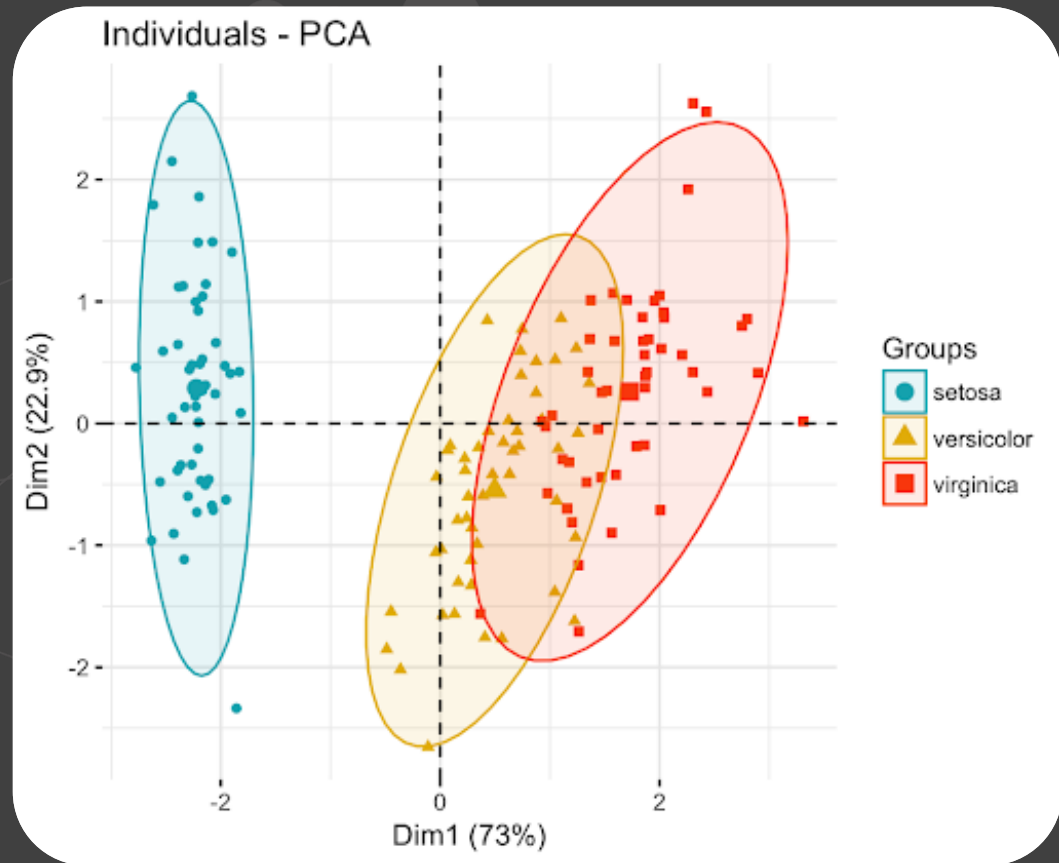
Normalização

- Garantir que as variáveis independentes possuem a mesma escala
- Mesmo efeito numérico na otimização independente da escala.



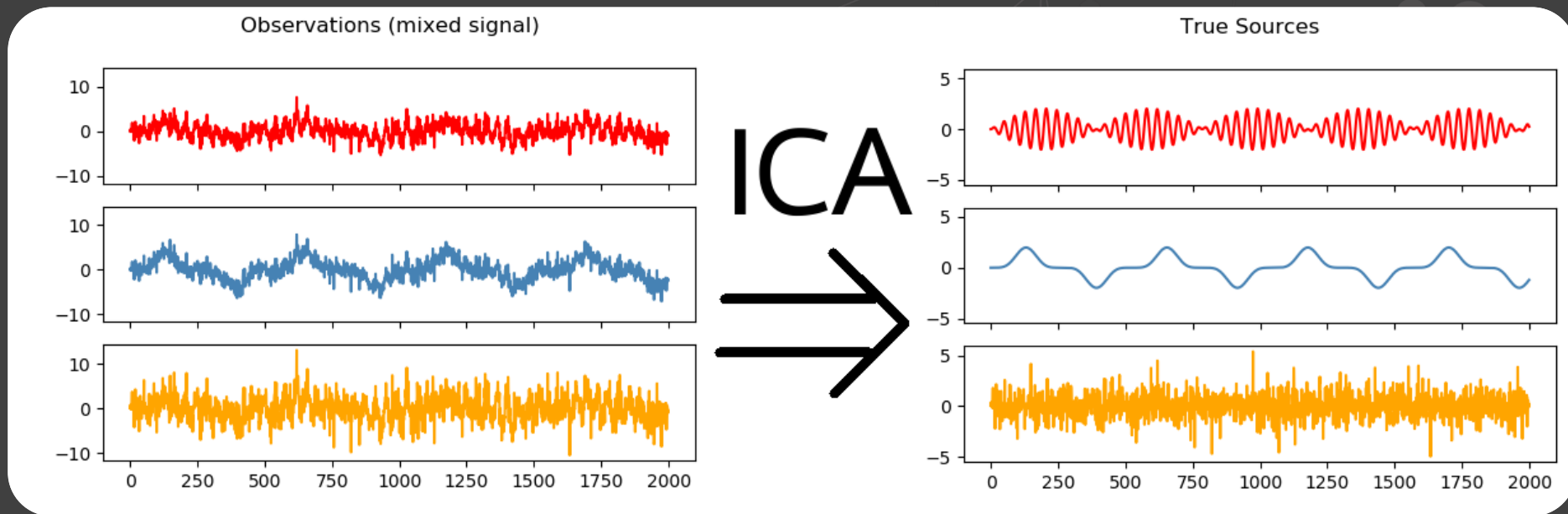
Análise de Componentes Principais

- Garantir que as variáveis independentes sejam descorrelacionadas.
- Identificar novas direções com maior concentração de energia / informação.



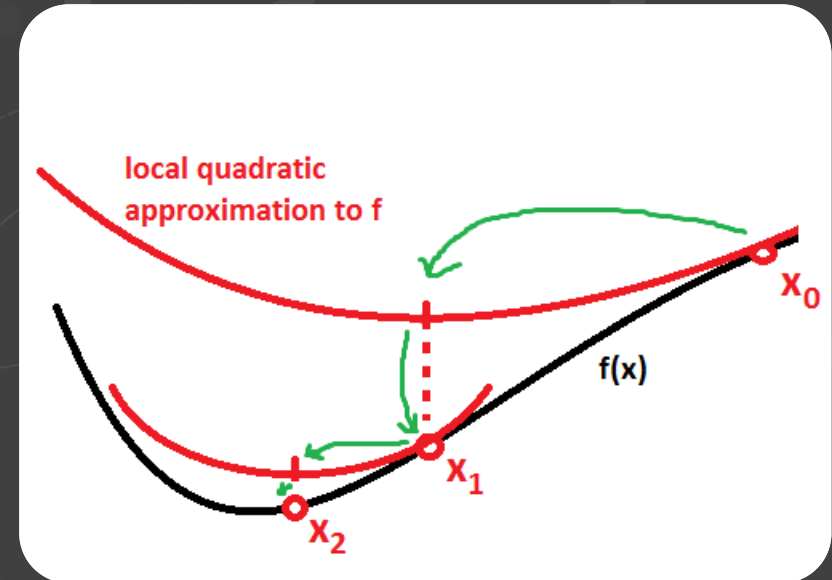
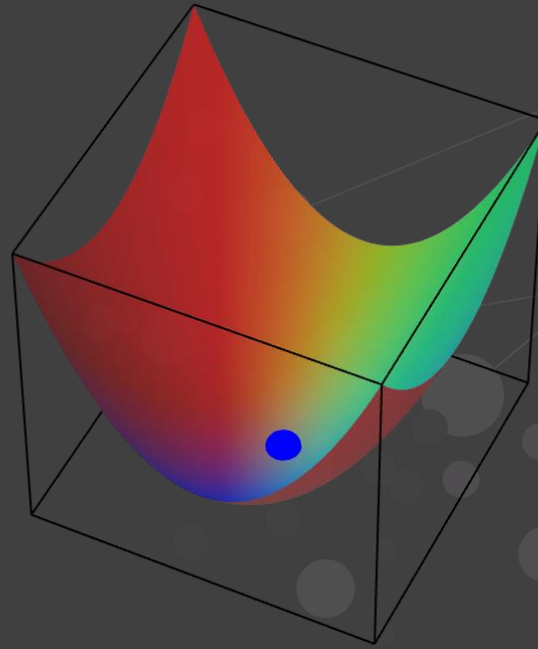
Análise de Componentes Independentes

- Garantir que as variáveis independentes sejam independentes
- Identificar principais direções de não-gaussianidade

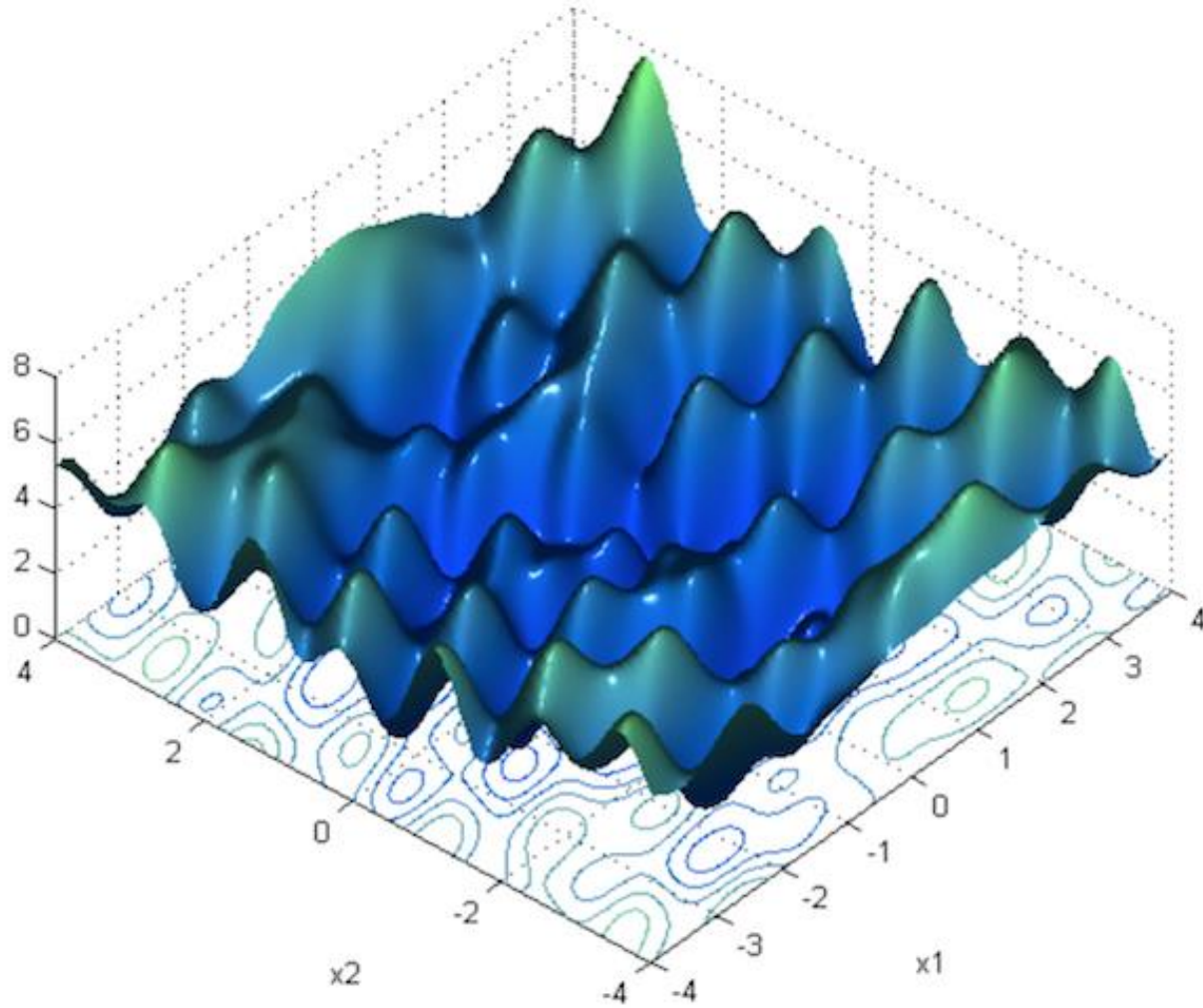


Encontrando o Ótimo Global para os Parâmetros

$$E = \frac{1}{N} \sum (y - f(w, x))$$



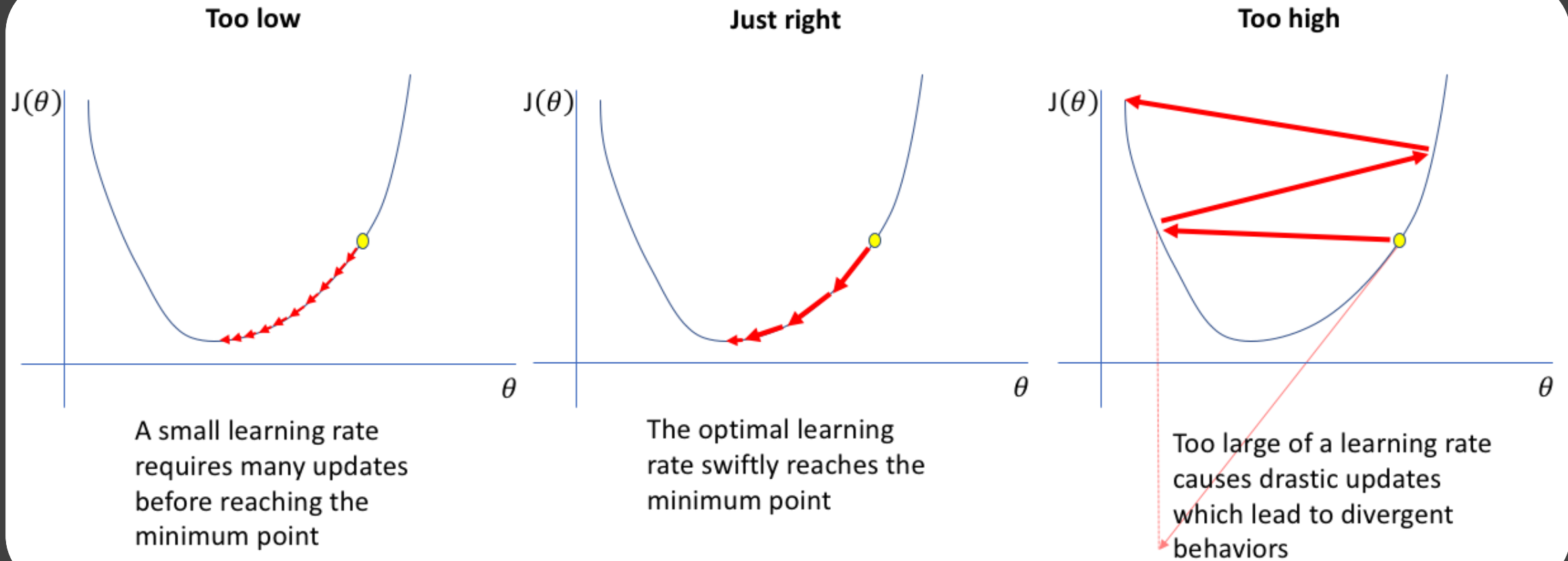
Wolfram Global Problem



Superfície do Erro
Médio Quadrático em
Função dos Parâmetros

$$E = \frac{1}{N} \sum (y - f(w, x))$$

Estimador de Gradiente + Taxa Correta



AdaGrad

**Decai a convergência em
proporção ao histórico de
atualização**

Parâmetros

- η : taxa de aprendizado
- ϵ : estabilizador do denominador

$$v_t^w = v_{t-1}^w + (\nabla w_t)^2$$

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{v_t^w + \epsilon}} * \nabla w_t$$

$$v_t^b = v_{t-1}^b + (\nabla b_t)^2$$

$$b_{t+1} = b_t - \frac{\eta}{\sqrt{v_t^b + \epsilon}} * \nabla b_t$$

RMSProp

**Permite selecionar a
intensidade do decaimento
com o hiperparâmetro β**

Parâmetros

- η : taxa de aprendizado
- ϵ : estabilizador do denominador
- β : Peso do gradiente atual no decaimento

$$v_t^w = \beta * v_{t-1}^w + (1 - \beta)(\nabla w_t)^2$$

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{v_t^w + \epsilon}} * \nabla w_t$$

$$v_t^b = \beta * v_{t-1}^b + (1 - \beta)(\nabla b_t)^2$$

$$b_{t+1} = b_t - \frac{\eta}{\sqrt{v_t^b + \epsilon}} * \nabla b_t$$

Adam

**Usa a média dos momentos
para controlar o decaimento.**

Parâmetros

- η : taxa de aprendizado
- ϵ : estabilizador do denominador
- β_1 : Peso do gradiente linear.
- β_2 : Peso do gradiente quadrático.

$$m_t = \beta_1 * m_{t-1} + (1 - \beta_1) * \nabla w_t$$

$$v_t = \beta_2 * v_{t-1} + (1 - \beta_2) * (\nabla w_t)^2$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} * \hat{m}_t$$

Quasi-Newton

Estima o 1º e o 2º momento
(gradiente e hessiana)

Goal:

$$\min_{x \in \mathbb{R}^n} f(x)$$

Gradient descent:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \alpha_k > 0$$

Newton method:

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

Modified Newton method: [Method of Deflected Gradients]

$$x_{k+1} = x_k - \alpha_k S_k \nabla f(x_k)$$

$$S_k \in \mathbb{R}^{n \times n}, \alpha_k \in \mathbb{R}$$

Special cases:

$$S_k = I_n: \text{Gradient descent}$$

$$S_k = [\nabla^2 f(x_k)]^{-1}: \text{Newton method}$$

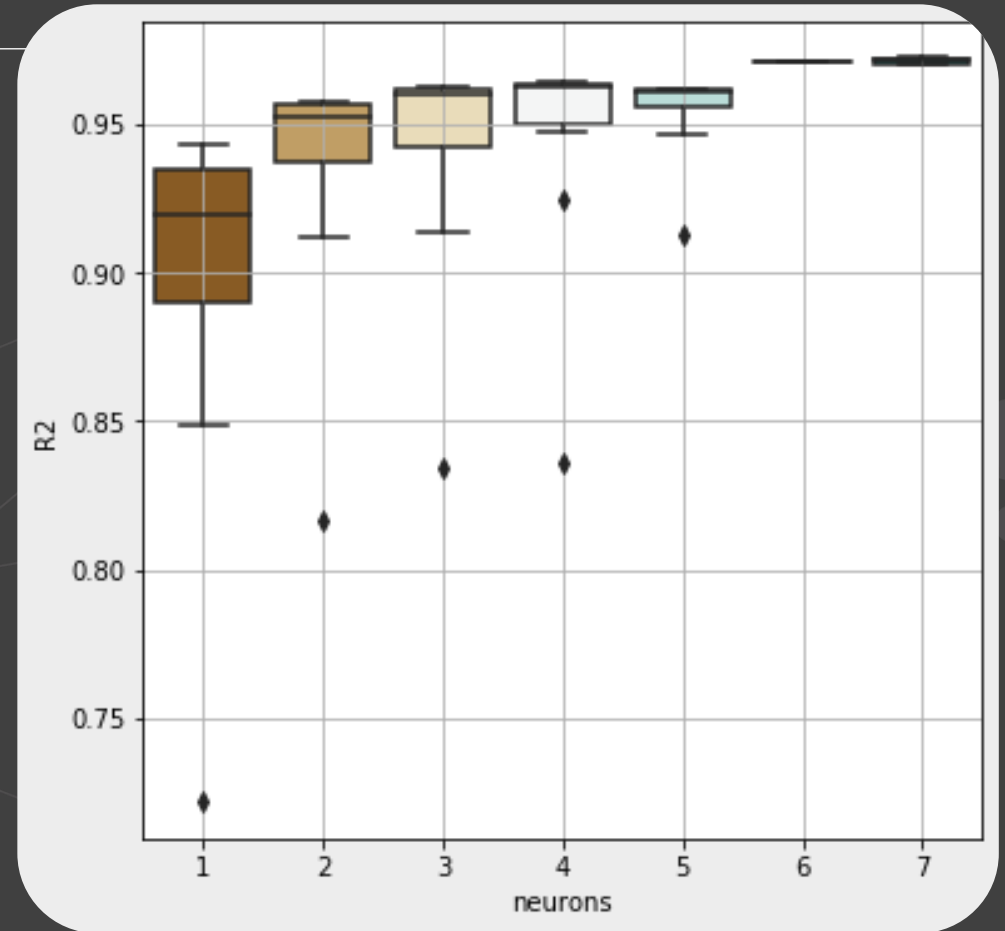
Annealing / Weight Decay

Simulated annealing algorithm

```
1  Select the best solution vector  $x_0$  to be optimized
2  Initialize the parameters: temperature  $T$ , Boltzmann's constant  $k$ , reduction factor  $c$ 
3  while termination criterion is not satisfied do
4      for number of new solution
5          for number of iterations
6              Select a new solution:  $x_0 + \Delta x$ 
7              if  $f(x_0 + \Delta x) > f(x_0)$  then
8                   $f_{\text{new}} = f(x_0 + \Delta x)$ ;  $x_0 = x_0 + \Delta x$ 
9              else
10                  $\Delta f = f(x_0 + \Delta x) - f(x_0)$ 
11                 random  $r(0, 1)$ 
12                 if  $r > \exp(-\Delta f / kT)$  then
13                      $f_{\text{new}} = f(x_0 + \Delta x)$ ,  $x_0 = x_0 + \Delta x$ 
14                 else
15                      $f_{\text{new}} = f(x_0)$ 
16                 end if
17             end if
18              $f = f_{\text{new}}$ 
19             Decrease the temperature periodically:  $T = c \times T$ 
20         end for
21     end while
```

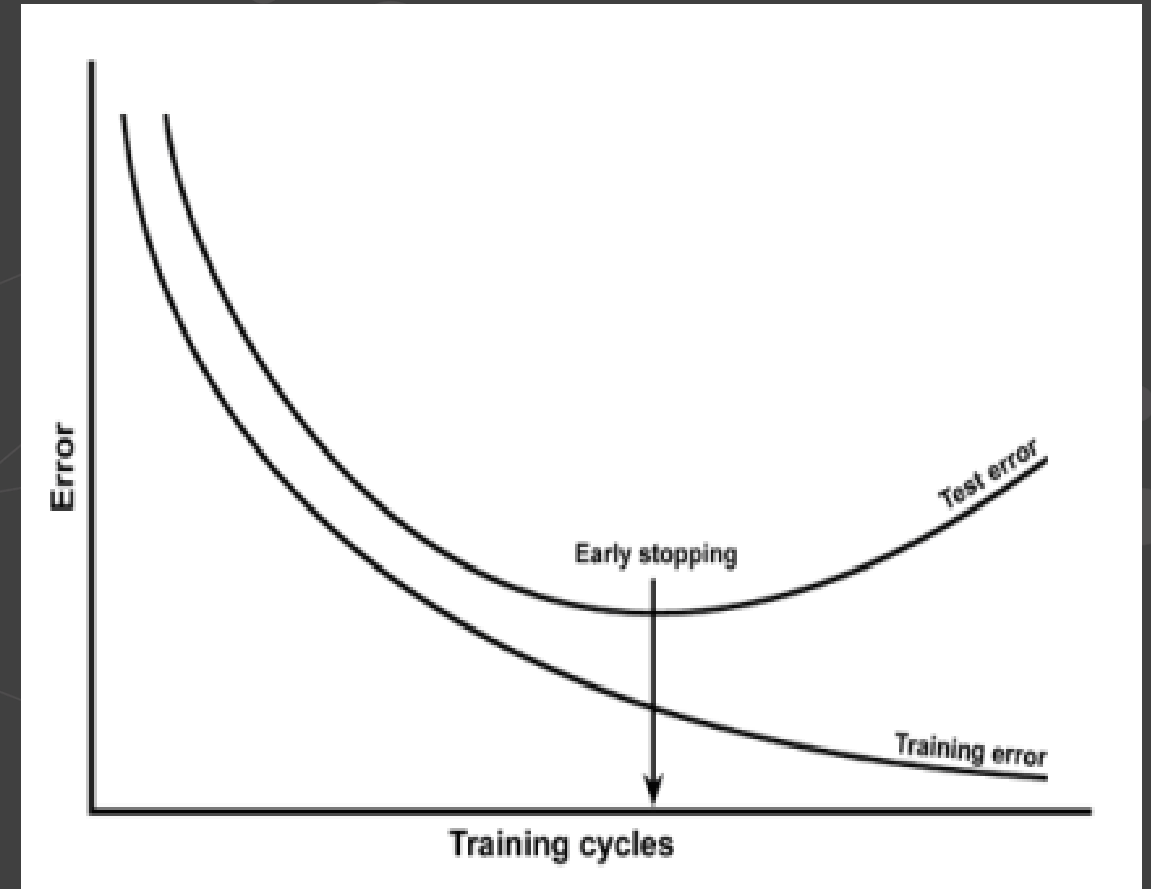
Validação

- Iteração dos hiperparâmetros
- Seleção da Figura de Mérito
- Seleção da Estatística de Ganho



Critérios de Parada

- Aumento no Erro de Validação
- Estabilidade da Figura de Mérito no Treino
- Estabilidade de Figura de mérito no Teste



Classificação Robusta Iris

Problema de Negócio

Características das flores

Largura & comprimento da pétala

Largura & comprimento da sépala



Iris Setosa



Iris Versicolor



Iris Virginica

Iris Setosa

Iris Versicolor

Iris Virginica

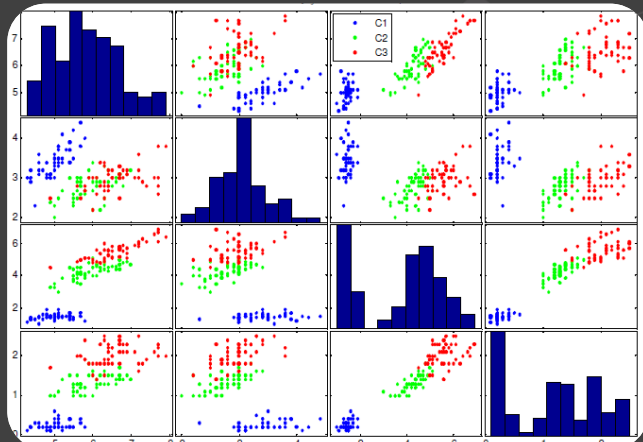
Representação



Características das flores

Largura & comprimento da pétala

Largura & comprimento da sépala



Espaço de
atributos com
4 dimensões!

Modelagem

Rede Neural Feed Forward

- Representação: 2 atributos
- Meta-parâmetros: 1..N neurônios TANH na camada oculta
- Treinamento: base de treino completa.
 - Precisão / Recall / Acurácia
 - Validação Cruzada 10 Folds
- Algoritmo RMSProp / Adam
 - RMSProp – Taxa de aprendizado fixa “Cautious”
 - Adam – Taxa de aprendizado com decaimento “Quickie”



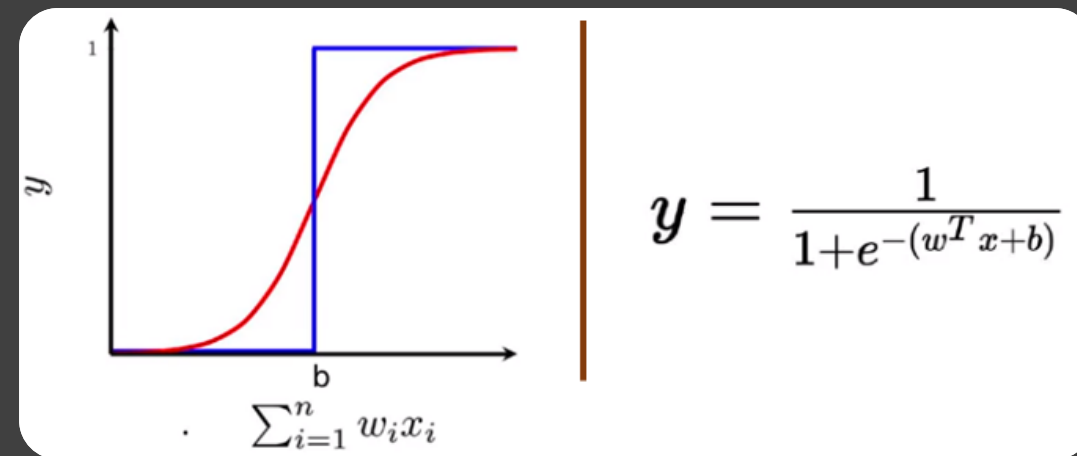
Iris Setosa



Iris Versicolor



Iris Virginica



Classificador Iris